Labour Force Survey-PLIDA linked microdata product

Guide to labour statistics

Explanatory notes providing information about the Labour Force Survey-PLIDA linked microdata product, and how to use it.

Released 18/03/2025

Sections

- Guide to labour statistics
  - Labour statistics recent and upcoming developments
  - About labour statistics
  - Earnings guide
  - Gender pay gap guide
  - Industry employment guide
  - Regional labour market data guide
  - Hours worked guide
  - Labour Force Survey products guide
    - Labour Force Survey concepts and data item guide
    - Labour Force Survey-PLIDA linked microdata product

On this page

- Introduction
- Scope and coverage
- Data items
- Scheduled updates
- Safe data treatments
- Usage tips and duplicate links
- Linkage and bias
- Limitations for longitudinal linkage
- References and resources
- Contact us

Close

Introduction

Official employment and unemployment estimates are available from Labour Force, Australia (/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release) . To complement these official estimates of employment and unemployment, ABS also produces a microdata product (available in ABS DataLab to approved

users) which links the Labour Force Survey (LFS) to the Person Level Integrated Data Asset (PLIDA). This linked LFS-PLIDA product, also known as the LFS PLIDA Modular Product (PMP), enables in-depth analysis of LFS data integrated with administrative datasets including health, education, government payments and taxation data. The LFS module in PLIDA contains monthly data from June 2023 onwards.

Given the nature of this linked dataset, it is not possible to replicate official employment and unemployment estimates, so data from the PMP should not be used to produce nor replicate official estimates. Any analysis or conclusions drawn from the linked LFS data should acknowledge the potential and, where possible, actual impact of bias on the interpretation of results. The data available in PLIDA are not market sensitive.

The monthly Labour Force Survey (LFS) provides information about the labour market activity of Australia's resident civilian population aged 15 years and over.

The LFS sample is designed and selected primarily to provide accurate estimates of employment and unemployment for the whole of Australia and, secondarily, for each state and territory. The ABS has been conducting the Labour Force Survey since 1960, initially as a quarterly survey. In February 1978, the frequency of the survey was changed from quarterly to monthly.

Households within selected dwellings are interviewed each month for eight months, with one-eighth of the sample being replaced each month. Information is obtained either by trained interviewers or through self-completion online. More detailed information about the LFS is available on the Labour Force, Australia methodology (/methodologies/labour-force-australia-methodology) page of the ABS website.

DataLab researchers may request access to the LFS module commencing January 2026.

Scope and coverage

The LFS surveys approximately 24,000 households each month which is equivalent to around 50,000 individuals.

The scope of the LFS is the resident civilian population of Australia aged 15 years and over, it excludes:

- Members of the permanent Australian defence forces
- Certain diplomatic personnel of overseas governments
- Overseas residents in Australia
- Members of non-Australian defence forces (and their dependents) stationed in Australia

The LFS uses administrative data in place of directly-collected responses for certain difficult to enumerate segments of the Australian population, consequently LFS data present in the PLIDA Modular Product also excludes populations living in:

- very remote areas
- non-private dwellings (e.g. prisons and care homes).

The LFS applies coverage rules to ensure that each person is associated with only one dwelling, and hence has only one chance of selection. The chance of a person being enumerated at two separate dwellings in the one survey is considered to be negligible.

The Labour Force Survey (LFS) is designed to survey the same household for eight consecutive months. However, individuals can move in and out of the LFS sample for several reasons, for example if they:

- do not complete the survey in a given month,
- are visiting the selected dwelling,
- move house permanently.

Data items

The LFS module consists of data obtained from the monthly survey. Survey month acts as a structural element. The monthly datasets can be joined to each other, and to the Spine, using linkage keys (see Usage Tips and Duplicate Links (#usage-tips-and-duplicate-links)).

Individual respondents to the LFS can be observed for up to eight months, making the data suitable for use in analysis of cross sections, pooled cross sections and short panels. Each month, approximately 6,000 individuals are added to the survey's sample as new rotation groups are enumerated.

On average, each cross section has approximately 48,000 observations. Each cross section is used separately to produce the headline LFS statistics. The unit record weights provided with the LFS module are those used to produce the published LFS statistics.

The weights in the LFS module do not aggregate to the population, and will not produce results comparable to published estimates due to the exclusion of out-of-coverage populations (i.e. residents of very remote areas and non-private dwellings).

This is further impacted, when interpreting estimates produced from linked records, due to the limitations inherent in linking data. No adjustments are made to the included weights to adjust for sample that cannot be linked. More information on the limitations and consequences of the linkage process is included in Usage Tips and Duplicate Links (#usage-tips-and-duplicate-links) , Linkage and Bias (#linkage-and-bias) , and Limitations for Longitudinal Linkage (#limitations-for-longitudinal-linkage) .

For more information about weighting and population benchmarking for the purposes of producing headline estimates, refer to the Labour Force, Australia methodology (/methodologies/labour-force-australia-methodology) page.

Estimates from the Labour Force Survey are predominantly based on individual records. However, hierarchical characteristics, such as items related to households and families, are attributed to each person record.

For variable and value descriptions, refer to the LFS PMP Data Item List.

LFS PMP Data Item List

Download XLSX
[49.69 KB]
(/statistics/understanding-statistics/guide-labour-statistics/labour-force-survey-products-and-data-item-guide/labour-force-survey-plida-linked-microdata-product/PLIDA%20Data%20Supply%20DIL%20-%20LFS.xlsx)

Scheduled updates

When new LFS data becomes available, the LFS module will undergo an incremental update. This occurs monthly, with a new month of data available in PLIDA around one week after the release of Labour Force, Australia, Detailed (/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/latest-release) . Unit weights for records in previous months' data will be revised each quarter consistent with the population benchmarking process for the LFS.

Safe data treatments

More information about safe data treatments is available within DataLab in the PLIDA Modular Product User Guide. However, the following steps were taken to treat the LFS data.

- The LFS person identifier is replaced with an anonymised PLIDA person ID SYNTHETIC_AEUID.

- The LFS household and dwelling identifiers HHID and DWELLID are anonymised.
- Detailed hours worked data items have been top coded to 99+ hours, this affects:
  - hours actually worked (HRAWMJ2)
  - hours usually worked (HRUWAJ99)
  - hours usually worked in main job (MUSLHRFX99)
  - hours actually worked in all jobs last week (WKDHOUR2)
  - number of hours preferred (underemployed) (PREFHOUR99)

Usage tips and duplicate links

General advice on working with PLIDA modules is available within DataLab in the PLIDA Modular Product User Guide.

ID Variables

The following table lists all ID variables present in the LFS module. These identifiers can be used to join information across monthly tables in the LFS module, and to identify relationships between LFS respondents. SYNTHETIC_AEUID can be used as the linkage key to link LFS to the PLIDA Spine via linkage files available in DataLab.

Key	Entity Type	Other PMP modules that use this key
SYNTHETIC_AEUID	LFS respondent identifier	N/A
HHID	Household identifier	N/A
DWELLID	Dwelling identifier	N/A

Records not linked to the Spine

When using the LFS module, you may encounter records which do not correspond to a Spine element (SPINE_ID = NULL). This can occur for the following reasons:

- Incomplete or inaccurate personal information was provided for the LFS respondent. For example, a respondent may decline to provide their first and/or last name, or their date of birth.
- Inconsistent information between the LFS and the Spine. For example, the respondent may have recently changed residential address.
- Absence from the Spine. The LFS respondent may not appear in the datasets used to form the Spine.

Duplicate links to the Spine

No statistical linkage process is perfect. In some rare instances, multiple SYNTHETIC_AEUIDs may be linked to a single SPINE_ID. Theoretically, this represents the linkage of multiple different LFS respondents to the same person represented on the Spine. However, in many of these instances a single LFS respondent is erroneously represented by multiple SYNTHETIC_AEUIDs.

In some (but not all) instances, apparent duplicate records are appropriate for use and should not necessarily be excluded from analysis. However, where apparent duplicate records cause inexplicable results, we advise removing all applicable records or using only the earliest instance of a record impacted by duplication (i.e. from the first survey month in which the record appears).

Broadly, the duplicate links can be categorised into two groups based on whether the root cause is on the survey side, or due to the PLIDA linkage method and/or spine creation.

The majority of the apparent duplicate links are a result of limitations with the LFS survey design, processing or responses. NB: These limitations affect only the quality of the linkage and not the published LFS estimates.

LFS-based duplicates

Duplicates may be observed when analysing a single month's LFS dataset for the following reasons:

1. Changes in household composition over time.

- Where an individual is not continuously in sample (i.e. they move into and out of a dwelling, or they change between being a visitor to and resident of a dwelling) they may be assigned a different person number within a household at different points in time. Where that occurs, they will be assigned multiple identifiers (which appear as SYNTHETIC_AEUIDs in PLIDA) over their time in sample.
- Additionally, a new resident or visitor in the dwelling may be assigned a vacant, but previously assigned, person number. This can result in the re-use of an LFS identifier over a dwelling's time-in-sample. After the first month-in, the LFS-PLIDA linkage method matches on the LFS identifier (i.e. the method assumes the persistent use of LFS identifiers). Subsequently, this can result in multiple different people being linked to the same Spine record within a given month.

2. Very rarely, errors in the way individual respondents are enumerated in the survey.

- Where a respondent completes the survey incorrectly it is possible for an individual to be enumerated more than once within the same month. This will appear as multiple SYNTHETIC_AEUIDs being linked to a single Spine entity).

Duplicate links may be observed when analysing multiple months of LFS datasets:

1. For the same reasons as for a single month. In particular, changes in household composition are more prevalent when analysing across multiple months than for any single month in isolation.
2. Additionally, apparent duplicate links may occur where the same person has continued to complete the survey after changing address, or an individual has been re-selected in the survey at a later date.

Where duplicate records are observed in the data for more than 8 months, and for discussion of issues specific to longitudinal analysis, please refer to Limitations for Longitudinal Linkage (#limitations-for-longitudinal-linkage).

PLIDA-based duplicates

Instances of duplicate links as a result of PLIDA processes are very rare with only a handful of examples across the 90,000+ unique records linked between LFS and the PLIDA Spine (as of December 2025).

Within a given month, they may occur:

- Where the linkage method identifies that two individual respondents are the same person. There are legitimate instances where this occurs but there is also error associated with this process, for example, twins living at the same address and who have very similar names may be erroneously linked to the same Spine entity.
- Due to Spine dataset imperfections, for example, where information from multiple unique people is intertwined in the formation of a single Spine entity resulting in links being found to multiple LFS respondents.

Across different time periods, duplicate links may exist:

- Where different LFS respondents have similar characteristics and consequently are each linked to the same single Spine entity.

Linkage and bias

Linkage

The LFS data from June 2023 to November 2023 are linked to the Spine using deterministic linkage methods. LFS data from December 2023 onwards are linked to the Spine using probabilistic linkage methods, see Person Linkage Spine (/about/data-services/data-integration/person-linkage-spine). Monthly linkage reports are available within DataLab and provide information on linkage rates and missingness rates for key linkage variables.

Any analysis or conclusions drawn from the linked LFS data should acknowledge the potential and, where possible, actual impact of bias on the interpretation of results.

Bias in linked data

When looking at aggregate statistics derived from linked observations it is important to be aware of bias and to attempt to control for it where possible. Separating bias from standard error is difficult, particularly for estimates from small samples where the variability inherent in the estimate can be large relative to the size of the estimate.

The Labour Household Surveys section of the ABS conducted comparative analysis of the labour market characteristics of the LFS sample pre and post linkage for data collected between June 2023 and June 2025. This analysis may be helpful in understanding some of the limitations of and biases present in the linked data.

For the unemployment rate, there is bias downwards for the linked group. This bias is broadly consistent across the characteristics chosen for analysis.

For the employment to population ratio, in aggregate there is no clear bias for the linked group. However, this is largely due to offsetting differences. For example, most age groups have a higher rate of employment in the linked group than the unlinked group, with a notable exception being the group aged 65 and over. There are also notable differences in the bias by sex, state of usual residence, relationship in the household and country of birth.

For the share of hours worked by people employed full-time, there is some similarity with the employment to population ratio. In aggregate there is no clear bias between the two groups, but this is largely due to offsetting differences present in many of the demographic subgroups, including: sex, country of birth, state of usual residence and age group.

Researchers should exercise caution when interpreting results from the linked LFS data and should not assume the linked data is representative of the full LFS sample. Re-weighting the linked LFS data may increase its representativeness, however, the ABS is not currently offering advice on re-weighting approaches for LFS data.

Unemployment rate

The LFS sample that has been successfully linked to the PLIDA Spine (referred to here as the linked group) is less unemployed than the LFS sample where no link to the PLIDA Spine has been found (referred to as the unlinked group), and is consequently less unemployed than the full LFS sample.

This means that any conclusions drawn from linked data will understate unemployment. In 24 of the 25 months, the unemployment rate from the linked group was lower than the unlinked group, with the difference being statistically significant for 16 of those 24 months.

This difference can be understood, in part, by demographic characteristics such as the age profile of linked respondents, for example:

The linked group has a lower share of 15-24 year old people than the unlinked group, and young people tend to have a higher rate of unemployment than other ages. Combined, this explains part of the observed difference. The lower linkage rates for young people may be a result of:

- greater mobility, resulting in geographical missingness/timing inconsistencies
- generational attitudes to sharing personal information, resulting in missing or inaccurate names, sex or ages and dates of birth
- delayed enrolment in administrative programs that form the PLIDA Spine.

Other demographic characteristics were also analysed to help understand the lower unemployment rate in the linked group:

- Sex: The difference in unemployment for women was more prevalent than for men: unlinked women were more unemployed than men relative to their respective linked populations, i.e. there's a more frequent bias in the unemployment rate for women than there is for men.

- State/territory: Significant differences between the linked and unlinked group were observed across states and territories. The more populous states showed the greatest number of significant differences, with all significant differences showing the linked group with lower unemployment than the unlinked group.
- Relationship in household: Household relationships did not tend to exhibit significant differences in the unemployment rate between the linked and unlinked groups, except in the case of more populous relationship types. In particular, those who were a spouse/partner or a child of someone else in the household were less unemployed in the linked group. This was most prevalent for spouses/partners.
  Conversely, the linked group was more unemployed than the unlinked group for visitors, unrelated individuals and male same-sex partners, but only when the groups were combined across the entire time period (individual periods had sample sizes too small to draw accurate conclusions).
- The following variables were considered but did not result in significant differences that were consistent across time periods:
  - year of arrival
  - country of birth.

Employment to population ratio

The rate of employment relative to the population, varies between the linked and unlinked group. Of the 25 periods analysed, the linked group had a significantly higher rate of employment than the unlinked group for 9 months and significantly lower for 7 months, with the remaining 9 months not exhibiting significant differences and being split evenly between the two groups. Note, having a lower unemployment rate in a population group does not necessarily result in a higher level of employment. The population also includes people not in the labour force.

Demographic characteristics were analysed to help understand the differences between the linked and unlinked groups. Some characteristics displayed differences between the two groups that were consistent across most time periods, while others varied across time.

Consistent differences:

- Age: The differences were significant for most age groups for every time period. Instances where there was no statistically significant difference included for the groups:
  - aged 65 or over, which displayed no significant difference between the linked and unlinked groups
  - aged 15-24 and 25-34 which were significant in most, but not all, months.
- Year of arrival: Overseas born populations in the linked group tended to have a higher rate of employment than those in the unlinked group. This was particularly true for more recent arrivals.

Variable differences:

- Sex: females in the linked group tended to have a higher rate of employment than females in the unlinked group, whereas males in the linked group tended to have a lower rate of employment.
- Relationship in household: People without children tended to have a lower rate of employment in the linked group, whereas people with children (and their children or dependents) tended to have a higher rate of employment in the linked group, relative to their respective unlinked cohorts.
  This result was true for both males and females, but the results were accentuated when cross-classified by sex. In other words, there were a greater number of significant differences for females when the linked group had a higher rate of employment for a particular family relationship, and a greater number of significant differences for males when the linked group had a lower rate of employment.
- State/territory:
  - Victoria, Queensland and the Northern Territory had significantly higher rates of employment in the linked group than the unlinked group
  - South Australia, Western Australia, Tasmania, and (for the most part) New South Wales, had significantly lower rates of employment in the linked group
- Country of birth:
  - The cohorts born in Australia and North-West Europe had significantly lower rates of employment in the

linked group than the unlinked group

- The cohorts born in Africa and Asia had significantly higher rates of employment in the linked group than the unlinked group
- The cohorts born in Southern and Eastern Europe, and the Americas had varying or few significant differences in the rates of employment between the linked and unlinked groups

Share of hours worked by full-time employees

The share of hours worked by full-time employees was very similar in the linked and unlinked groups when considered in aggregate. Only two periods had a significant difference between the two groups (both slightly higher in the linked group). However, examining the differences by various demographics exhibited some stark differences between the linked and unlinked groups.

These results illustrate that care should be taken when analysing estimates related to hours worked. They also illustrate that results drawn from an aggregate population cannot necessarily be generalised across its component subpopulations.

- Sex: females in the linked group had a smaller share of hours worked by full-time workers than females in the unlinked group, whereas males in the linked group had a larger share of hours worked by full-time workers.
- Country of birth: the most consistent differences were evident for the cohort born in Australia offsetting that born in Southern and Central Asia. The Australian-born cohort had a smaller share of hours worked by full-time workers in the linked group than in the unlinked group, whereas the cohort born in Southern and Central Asia had the opposite: a larger share of hours worked by full-time workers in the linked group compared to the unlinked. The cohorts born in South-East Asia and the Americas had similar results to that born in Southern and Central, but the results were not as consistently significant across time.
- Year of arrival: the results were very similar to those observed for country of birth. The cohort of migrants arriving 1 to 2 years ago had a larger share of hours worked by full-time workers in the linked group compared to the unlinked. This was offset by the opposite relationship between the linked and unlinked group being observed for the Australian-born population, as noted above.
- State/territory: there were only a small number of statistically significant results when looking at state or territory. For New South Wales, there were 6 months with significant differences between the linked and unlinked groups, with each exhibiting a larger share of hours worked by full-time workers in the linked group. Tasmania exhibited the opposite, with a smaller share of hours worked by full-time workers in the linked group, with significant differences in 5 months.
- Age: there were significant differences observed by age group, predominantly for the cohorts aged:
  - between 25 and 34 years (larger share of hours worked by full-time workers in the linked group compared to the unlinked group)
  - 65 years or older (smaller share of hours worked by full-time workers in the linked group compared to the unlinked group).
- Relationship in household: the differences were rarely significant. Where there were significant differences, they were consistent within a cohort. They predominantly related to those who were:
  - living in group households (larger share of hours worked by full-time workers in the linked group compared to the unlinked group)
  - dependent students or those living alone (smaller share of hours worked by full-time workers in the linked group compared to the unlinked group).

Summary Table

Demographic Characteristic	Unemployment Rate	Employment to Population Ratio	Share of hours worked by people employed full-time
Overall	Linked group is less unemployed than linked group.	Direction of difference between linked and unlinked groups employment is inconsistent across time.	No significant differences in aggregate: offsetting differences by characteristic (see below)
Sex	Linked women less unemployed than unlinked women to a greater extent than linked men are less unemployed than unlinked men.	Linked women more employed than unlinked women*. Linked men less employed than unlinked men*.	Linked women: lesser share of hours worked by full-time workers. Linked men: greater share of hours worked by full-time workers.
Age	Linked group is less unemployed: fewer young people in linked group.	Linked group is more likely to be employed than the unlinked group for every age except 65+ years old.	Linked cohort aged 65+: lesser share of hours worked by full-time workers. Linked cohort aged 25-34: greater share of hours worked by full-time workers.
State and territory	Linked group was less unemployed than unlinked irrespective of geography.	Linked Vic, QLD and NT groups were more employed than unlinked groups*. Linked NSW, SA, WA and Tas groups were less employed than the unlinked group*.	Linked Tasmania cohort: lesser share of hours worked by full-time workers. Linked NSW cohort: greater share of hours worked by full-time workers.
Country of birth	No significant differences	Linked overseas-born people more likely to be employed than unlinked overseas-born group.	Linked cohort born in Australia: lesser share of hours worked by full-time workers. Linked cohort born in Southern and Central Asia: larger share of hours worked by full-time workers.
Relationship in household	Linked people with partner/ spouse in household less unemployed than unlinked group.	Linked people with children tend to be more employed than unlinked group*. Linked people with children tend to be less employed than unlinked group*.	Some significant differences for small cohorts: people living group households, dependent students, and people living alone.

* Strength of bias (significance of difference) is variable across time.

Further action

It is clear, from the results outlined above, that there are many statistically significant differences between the linked and unlinked LFS sample. These differences will result in bias when producing estimates from the linked sample.

Researchers may wish to consider re-weighting the linked records to account for bias. However, the ABS is not currently offering advice on re-weighting approaches for LFS data.

As noted previously, any analysis or conclusions drawn from the linked LFS data should acknowledge the potential and, where possible, actual impact of bias on the interpretation of results.

Limitations for longitudinal linkage

For general information about the longitudinal aspect of Labour Force Survey data, and some of its limitations, please refer to Microdata: Longitudinal Labour Force, Australia (/statistics/microdata-tablebuilder/available-microdata-tablebuilder/longitudinal-labour-force-australia) .

Along with the duplicates and bias elements, there are additional considerations which apply when analysing the linked LFS data longitudinally. In the linked LFS data there are some instances where individual people appear to be

linked to the Spine for more than 8 survey months. In most instances this is a result of people with similar linkage characteristics (name, age, sex), who have been selected in the LFS sample within a similar geographical area as those in previously selected dwellings, being linked to a single Spine entity. These are different people being linked to the same Spine entity, at different points in time, due to the quality of their linkage information combined with the linking method.

There may also be rare instances of the same person completing the LFS over more than 8 months, as a result of respondents changing residence and their new dwelling being selected in a new or existing rotation group.

References and resources

- DataLab access (/statistics/microdata-tablebuilder/datalab)
- Person Level Integrated Data Asset (PLIDA) (/about/data-services/data-integration/integrated-data/person-level-integrated-data-asset-plida)
- Labour Force, Australia (/statistics/labour/employment-and-unemployment/labour-force-australia/latest-release) (latest release)
- Labour Force, Australia, Detailed (/statistics/labour/employment-and-unemployment/labour-force-australia-detailed/latest-release) (latest release)
- Labour Force, Australia Methodology (/methodologies/labour-force-australia-methodology)
- Labour Statistics: Concepts, Sources and Methods (/statistics/detailed-methodology-information/concepts-sources-methods/labour-statistics-concepts-sources-and-methods/latest-release)
- Guide to labour statistics (/statistics/understanding-statistics/guide-labour-statistics)

Contact us

Please get in contact with us via labour.statistics@abs.gov.au (mailto:labour.statistics@abs.gov.au) if you have any queries or feedback relating to this guide.